Twitter's Big Hitters

Peter Laflin
Bloom Agency
Leeds, UK
plaflin@bloomagency.co.uk

Alexander V. Mantzaris
Department of Mathematics and Statistics
University of Strathclyde, UK
alexander.mantzaris@strath.ac.uk

Desmond J. Higham
Department of Mathematics and Statistics
University of Strathclyde, UK
d.j.higham@strath.ac.uk

Peter Grindrod
Department of Mathematics
University of Reading, UK
p.grindrod@reading.ac.uk

Fiona Ainley
Bloom Agency
Leeds, UK
fainley@bloomagency.co.uk

Amanda Otley
Bloom Agency
Leeds, UK
aotley@bloomagency.co.uk

ABSTRACT

We describe the results of a new computational experiment on Twitter data. By listening to Tweets on a selected topic, we generate a dynamic social interaction network. We then apply a recently proposed dynamic network analysis algorithm that ranks Tweeters according to their ability to broadcast information. In particular, we study the evolution of importance rankings over time. Our presentation will also describe the outcome of an experiment where results from automated ranking algorithms are compared with the views of social media experts.

Categories and Subject Descriptors

J.4 [**SOCIAL AND BEHAVIORAL SCIENCES**]: Sociology—*communication network measures*

General Terms

MANAGEMENT, HUMAN FACTORS, MEASUREMENT, THEORY

Keywords

centrality, dynamic network, social network analysis, Twitter

1. MOTIVATION

The success of many commercial organisations has become very closely tied to the level and quality of their penetration across online social media. Issues such as

- how is my brand performing?
- how successful was that advertising campaign?
- how big a problem was that recent indiscretion from our celebrity endorsee?
- what do the public like/dislike about the film that we premiered last weekend?
- what will be the next big Christmas toy?

can be **investigated** by analysing the digital trails left by human social interaction. More importantly a key strategy for **controling** these issues is to interfere appropriately within the online digital world. To do this, it is important to identify **influential** players.

We focus here on the analysis of Twitter data as a tool to drive business analytics. Our approach is to listen to relevant, topic-based, Twitter activity in order to build a dynamic network that summarizes the information flow between users. In this way, by generalizing standard centrality measures developed for static networks, we can discover which players are currently the most influential in relation to the topic.

2. SET UP

Recent network-style analysis of Twitter activity has included the work in [1], which focuses on transmission of shortened URLs, [8], which looks at global flow of information, and [2, 6], which look at rankings over the follower network. The experiment that we describe differs in that we focus on subject-specific Tweets that would be of interest in a typical business application and therefore analyse a subset of Tweeters relative to this topic who make up what we regard as an appropriate 'active' subnetwork.

Having generated a time-stamped sequence of interactions (who Tweeted, when, and who follows these Tweeters), we may then use the dynamic network centrality measures from [3] to discover the important players.

3. DATA

In order to model a typical client-driven study, we constructed a dynamic network based on the phrases `city break`, `cheap holiday`, `travel insurance`, `cheap flight` and two brand-specific names. The clock ran over the 22 hour period from 14:41 on 17th June, 2012 to 12:41 on 18th June, 2012. We extracted a list of 590 active nodes and summarized the

DE 2012 Aberdeen October 23rd - 25th 2012

	first	second	third	fourth	fifth
b=0	74	397	361	34	357
b=0.2	397	361	357	396	74
b=0.5	397	361	357	74	374
b=2	581	464	414	552	505

Table 1: Eventual top five nodes for each value of time-downweighting parameter, b. (These are the Tweeters appearing in Figure 1.)

Twitter activity into 20 time windows, each covering a 66 minute period. This is precisely the type of evolving connectivity data that can be analysed with the dynamic broadcast centrality measure in [3].

Building on the standard Katz centrality for static networks [5, 10] and the dynamic extension in [4], the broadcast measure in [3] computes a running "communicability" score between every pair of nodes (Tweeters) by counting how many different routes (more precisely, dynamic walks) through the network have currently been made available. The measure involves two parameters:

downweighting for length: a parameter $0 < a < 1$ is chosen so that walks using w edges are downweighted by a^w,

downweighting for time: a parameter $b \geq 0$ is chosen so that walks beginning t time units ago are downweighted by e^{-bt}.

Downweigting in this manner reflects the idea that

- information may become corrupted as it passes between players,
- information may lose relevance over time.

Overall, by measuring the total ability of a node to communicate over time to all other nodes in the network, we arrive at broadcast scores that can be used for ranking. Nodes with high broadcast scores can therefore be viewed as highly influential players. We emphasize that broadcast ability in this time-dependent sense takes account of knock-on effects; for example, where a more important message may be more quickly passed on and more widely dispersed by its recipients. This type of time-dependent phenomenon is not picked up by applying static network measures or focusing on overall bandwidth [9].

4. EXPERIMENTAL RESULTS

For illustrative purposes, we present here some results based on four different choices of b, the time downweighting parameter. We use values of 0, 0.2, 0.5 and 2.0. In each case, we focus on the five Twitter accounts that are top-ranked at the final time point. Figure 1 shows how the dynamic, time-sensitive rankings of these 'eventual big hitters' varies over time.

Table 1 shows which of the 590 nodes form the top five in each case. (These node labels are local to the experiment and have no further significance.)

We note that the case $b = 0$ corresponds to no downweighting over time—interactions that began several hours ago are given the same weight as very recent conversations. As b increases, an ageing effect kicks in, with $b = 2$ being the

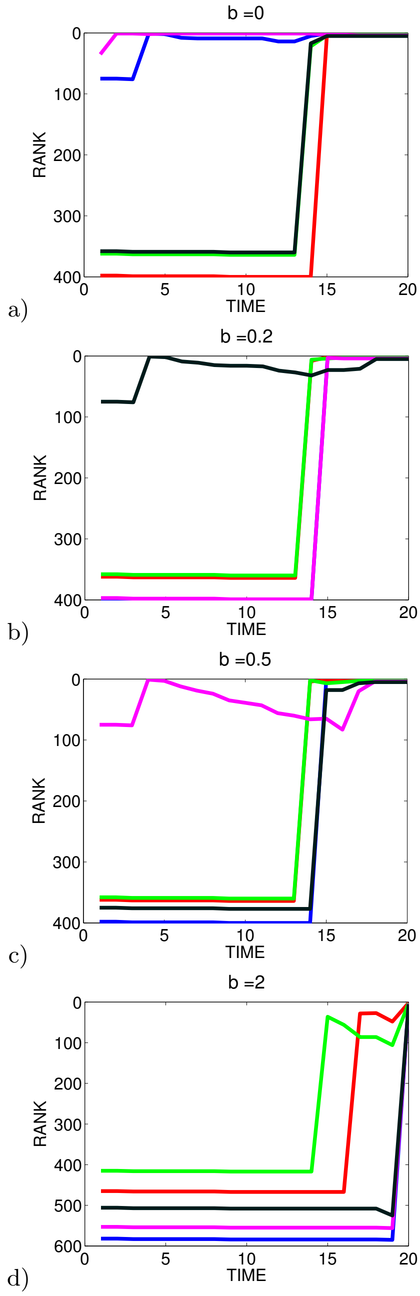

Figure 1: Journeys to the top of the rankings. Subfigures a) to d) have time-downweighting parameter b equal to 0, 0.2, 0.5, and 2.0, respectively. In each case, we focus on the five Tweeters who become top ranked at the final time point. The plots show the evolution of their rankings over time.

closest to a static ‘snapshot’ view, where we only consider Tweets that happened in the current time window.

In Table 1 we see consistency between the final-time rankings for b equal to 0, 0.2 and 0.5—four of the five nodes are common to all three rows. However, when b is increased to 2 the top five list changes completely. If we regard the final time point as our key target date, we see from Figure 1 that the choices of b equal to 0, 0.2 and 0.5 also identify their key nodes more quickly (at around the 15th time window) than the more snapshot-based $b = 2$ version. These conclusions agree with other tests in [3] that focused on centrality prediction with email data, where making some use of historical interaction, rather than focusing entirely on current activity, was found to be effective.

5. FOLLOW-UP ANALYSIS

In the corresponding presentation that accompanies this write-up, we will also discuss separate analytical results on this Twitter network [7], including

- interpreting the dynamic communicability ranking scores in relation to semantic information concerning the Twitter accounts,
- discussing the behaviour of accounts that generate automated Tweets,
- comparing computer-generated importance rankings against those of social media experts.

Acknowledgement This work was performed within the *Mathematics of Large Technological Evolving Networks* (MOLTEN) project, which is supported by the Engineering and Physical Sciences Research Council and the Research Councils UK Digital Economy programme, with grant ref. EP/I016058/1.

6. REFERENCES

[1] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, *Everyone’s an influencer: quantifying influence on Twitter*, in Proceedings of the fourth ACM international conference on Web search and data mining, WSDM ’11, New York, NY, USA, 2011, ACM, pp. 65–74.

[2] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, *Measuring user influence in Twitter: The million follower fallacy*, in ICWSM 10: Proceedings of international AAAI Conference on Weblogs and Social, 2010.

[3] P. Grindrod and D. J. Higham, *A matrix iteration for dynamic network summaries*, SIAM Review, Research Spotlights, to appear, (2012).

[4] P. Grindrod, D. J. Higham, M. C. Parsons, and E. Estrada, *Communicability across evolving networks*, Physical Review E, 83 (2011).

[5] L. Katz, *A new index derived from sociometric data analysis*, Psychometrika, 18 (1953), pp. 39–43.

[6] H. Kwak, C. Lee, H. Park, and S. Moon, *What is Twitter, a social network or a news media?*, in Proceedings of the 19th international conference on World wide web, WWW ’10, New York, NY, USA, 2010, ACM, pp. 591–600.

[7] P. Laflin, A. V. Mantzaris, F. Ainley, A. Otley, P. Grindrod, and D. J. Higham, *Dynamic targeting in an online social medium*, Tech. Rep. 16, University of Stratchlyde, September 2012. (To appear in SocInfo 2012).

[8] K. Lerman, R. Ghosh, and T. Surachawala, *Social contagion: An empirical study of information spread on Digg and Twitter follower graphs*, CoRR, abs/1202.3162 (2012).

[9] A. V. Mantzaris and D. J. Higham, *A model for dynamic communicators*, European Journal of Applied Mathematics, to appear, (2012).

[10] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, Cambridge, 1994.